E C O N O M I C S   B U L L E T I N

A note on the relationship between the information matrx test and a score test for parameter constancy

Daisuke Nagakura
Department of Economics, University of Washington

Abstract

Information matrix (IM) test (White, 1982) has been used for detecting general model misspecification in the applied econometrics literature. Two of the most commonly used asymptotic covariance matrix estimators (ACMEs) for the IM test are the one that White (1982) proposed in his original paper and Chesher (1983)'s ACME. Chesher (1984) showed that the IM test is in effect a score test for parameter constancy. In this note, I show that the IM test with White's ACME is not only the score test but also a specification robust form of the score test or a score test for quasi-maximum likelihood estimators. Based on this result, it is argued that we should be careful in selecting the ACME for properly interpreting the consequence of the IM test.

I wish to thank Charles Curry for his many helpful comments on the preliminary draft and kind encouragement. Earlier versions of the note were presented under different titles at the University of Washington, and the Pennsylvania Economic Association 2006 Annual Conference. I would like to thank Kum-Hwa Oh, Eric Zivot, and Richard Startzs, as well as the participants for their useful comments. Any remaining errors are mine.
Citation: Nagakura, Daisuke, (2008) "A note on the relationship between the information matrx test and a score test for parameter constancy." *Economics Bulletin,* Vol. 3, No. 5 pp. 1-7
Submitted: July 21, 2007. **Accepted:** January 21, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-07C10004A.pdf

1 Introduction

Information matrix (IM) test (White, 1982, 1983) has been used for detecting general model misspecification in the applied econometrics literature. The IM test exploits the well-known "information matrix equality," which states that if a model is correctly specified, the expectation of the sum of the Hessian matrix and the outer product of the gradient of a contribution of the log-likelihood is zero. The IM test rejects the null hypothesis that the model is correctly specified when the sample average of the sum is significantly different from zero. Chesher (1984) showed that the sum is equivalent to a score function for testing parameter constancy against certain random parameter models. Based on the result, Chesher (1983) proposed using the outer product of the gradient (OPG) estimator as an asymptotic covariance matrix estimator (ACME) for the IM test (see also Lancaster, 1984), which is one of the two most commonly used ACMEs; another one is the one that White (1982) proposed in his original paper. Many papers have attempted to improve the finite sample performance of the IM test by using different ACMEs as well as applying simulation methods. See, for example, Orme (1990), Horowitz (1994), and Dhaene and Hoorelbeke (2004) among others.

In this note, I develop Chesher (1984)'s score test interpretation of the IM test. It is shown that the IM test "with White's ACME" is not only a score test for parameter constancy, but also a specification robust form of the score test or a score test for quasi-maximum likelihood estimators (QMLEs) (White, 1982, p.8).[1] As shown in White (1982, pp.3-4), the QMLE is strongly consistent for the parameter vector which minimizes the Kullback-Leibler Information Criterion (KLIC). The result implies that the IM test with White's ACME rejects the null when the KLIC of a constrained model is significantly higher than the KLIC of the unconstrained model. The result also indicates that the IM test with White's ACME is powerless against misspecification in that we cannot lower the KLIC by incorporating random parameter fluctuations.

2 Information Matrix Test and Score Test

2.1 Information Matrix Test

Suppose that $\mathbf{y}_1, ..., \mathbf{y}_n$ are n observations from i.i.d. random vectors $\mathbf{Y}_1, ..., \mathbf{Y}_n$. Consider a problem to fit a model to the data using a family of parametric pdf or pmf (p.f.) $g(\mathbf{y}_i|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^T$ is a p - vector of parameters. The log-likelihood function is defined as $l(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$, where $l_i(\boldsymbol{\theta}) \equiv \log g(\mathbf{y}_i|\boldsymbol{\theta})$. Hereafter, summations are always taken from $i = 1$ to n and so the limits are omitted. Let $\mathbf{d}_i(\boldsymbol{\theta}) \equiv \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, and $\mathbf{F}_i(\boldsymbol{\theta}) \equiv \frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$. Hereafter, the argument $\boldsymbol{\theta}$ is suppressed where there is no ambiguity. For example, $\mathbf{F}_i(\boldsymbol{\theta})$ is abbreviated to $\mathbf{F}_i$. I use capital letters for matrices and small letters for vectors. A bold font is used for matrices and vectors.

It is well-known that if $g(\mathbf{y}_i|\boldsymbol{\theta})$ is correctly specified, i.e., the true model is given by $g(\mathbf{y}_i|\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the parameter space of $\boldsymbol{\theta}$, then $E[\mathbf{D}_i(\boldsymbol{\theta}_0) + \mathbf{F}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$, where $\mathbf{D}_i \equiv \mathbf{d}_i \mathbf{d}_i^T$ and $\mathbf{d}_i^T$ denotes the transpose of $\mathbf{d}_i$. This equality is known as "the information matrix equality." Here, and in what follows, expectations are taken with respect to the true distribution. The IM test proposed by White (1982) is a specification test that detects any general model misspecification violating this equality. It is based on a sample analogue of $E(\mathbf{D}_i + \mathbf{F}_i)$, namely, $n^{-1} \sum \mathbf{s}_i$, where $\mathbf{s}_i \equiv vech(\mathbf{D}_i + \mathbf{F}_i)$. Here, $vech(.)$

[1] Actually, this result is implicit in White (1982) and Chesher (1984)'s results since White's ACME was constructed so that it is consistent even when the model is misspecified. However, the result has not been proved rigorously in the literature.

is an operator stacking up only different elements in a symmetric matrix. For example, for a 3×3 symmetric matrix $\mathbf{A} = [a_{ij}]$, we have $vech(\mathbf{A}) = (a_{11}, a_{21}, a_{31}, a_{22}, a_{23}, a_{33})^T$ (see Magnus and Neudecker, 1999, p.49). Obviously, if $g(\mathbf{y}_i|\boldsymbol{\theta})$ is correctly specified, $E[\mathbf{s}_i] = \mathbf{0}$. Let $\mathbf{s} \equiv \sum \mathbf{s}_i$, $\mathbf{G}_i \equiv \frac{\partial \mathbf{s}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$ and $\mathbf{V} \equiv E(\mathbf{v}_i\mathbf{v}_i^T)$, where $\mathbf{v}_i \equiv \mathbf{s}_i - E(\mathbf{G}_i)[E(\mathbf{F}_i)]^{-1}\mathbf{d}_i$. Under the null hypothesis that the model is correctly specified and several regularity conditions, White (1982, Theorem 4.1, p.11) showed that $n^{-1/2}\mathbf{s}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}_0))$, where $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimate (MLE) of the unknown parameter vector $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$ (throughout the note, I assume that the regularity conditions in White, 1982, are satisfied). An ACME proposed by White (1982) is

$$\begin{aligned} \mathbf{V}_W \equiv & \ n^{-1} \sum [\mathbf{s}_i - (n^{-1}\mathbf{G})(n^{-1}\mathbf{F})^{-1}\mathbf{d}_i] [\mathbf{s}_i - (n^{-1}\mathbf{G})(n^{-1}\mathbf{F})^{-1}\mathbf{d}_i]^T \\ = & \ n^{-1} \left(\mathbf{S} - \mathbf{G}\mathbf{F}^{-1}\mathbf{C} - \mathbf{C}^T\mathbf{F}^{-1}\mathbf{G}^T + \mathbf{G}\mathbf{F}^{-1}\mathbf{D}\mathbf{F}^{-1}\mathbf{G}^T \right), \end{aligned}$$

where $\mathbf{D} \equiv \sum \mathbf{D}_i$, $\mathbf{F} \equiv \sum \mathbf{F}_i$, $\mathbf{G} \equiv \sum \mathbf{G}_i$, $\mathbf{S} \equiv \sum \mathbf{S}_i$, $\mathbf{C} \equiv \sum \mathbf{C}_i$, $\mathbf{S}_i \equiv \mathbf{s}_i\mathbf{s}_i^T$, and $\mathbf{C}_i \equiv \mathbf{d}_i\mathbf{s}_i^T$. Let $\mathbf{Q}_W \equiv \mathbf{S} - \mathbf{G}\mathbf{F}^{-1}\mathbf{C} - \mathbf{C}^T\mathbf{F}^{-1}\mathbf{G}^T + \mathbf{G}\mathbf{F}^{-1}\mathbf{D}\mathbf{F}^{-1}\mathbf{G}^T$. The IM test is defined as:

$$\text{IM} \equiv n^{-1}\mathbf{s}(\widehat{\boldsymbol{\theta}})^T\mathbf{V}_W^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{s}(\widehat{\boldsymbol{\theta}}) \quad \text{or} \quad \mathbf{s}(\widehat{\boldsymbol{\theta}})^T\mathbf{Q}_W^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{s}(\widehat{\boldsymbol{\theta}}), \tag{1}$$

which is shown to be distributed asymptotically as $\chi^2(q)$, i.e., the chi-square distribution with q degrees of freedom when the model is correctly specified, where $q \equiv p(p+1)/2$ is the number of elements of $\mathbf{s}_i$.

2.2 Score Test for Parameter Constancy

Consider the same model in Section 2.1 except that $\boldsymbol{\theta}$ is replaced by an i.i.d. p-variate random parameter vector $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{ip})^T$. Assuming this sort of random parameter fluctuation over i is a convenient way to incorporate individual heterogeneity. The conditional pdf of $\mathbf{Y}_i$ is denoted by $g(\mathbf{y}_i|\boldsymbol{\beta}_i)$. I assume the following two assumptions:

Assumption 1: The distribution of $\boldsymbol{\beta}_i$ is the p-variate normal with mean vector $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Omega} = [\omega_{ij}]$, and thus the pdf is $f(\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Omega}) \equiv (2\pi)^{-p/2}|\boldsymbol{\Omega}|^{-1/2}\exp[-(\boldsymbol{\beta}_i - \boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\theta})/2]$.

Assumption 2: For each $\mathbf{y}_i$, the function $g(\mathbf{y}_i|\boldsymbol{\beta}_i)$ of $\boldsymbol{\beta}_i$ is continuous at $\boldsymbol{\beta}_i = \boldsymbol{\theta}$ and its partial derivatives with respect to β_{ik} $k = 1, ..., p$ exist to fourth order in a p-ball $B(\boldsymbol{\theta})$, and these partial derivatives are continuous at $\boldsymbol{\beta}_i = \boldsymbol{\theta}$.

Assumption 2 is the smoothness assumption for applying the main tool in this note, the *Dirac delta function*, which is obtained as the limit of $f(\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Omega})$ for which $\boldsymbol{\Omega} \to \mathbf{0}$.[2] Assumption 1 can be relaxed so that $\boldsymbol{\beta}_i$ is an unknown function of a multivariate normal random vector since the resulting score test statistics take the same form as long as the functions are continuous and monotonically increasing.[3]

Let the unknown (fixed) model parameters be $\boldsymbol{\psi} \equiv (\boldsymbol{\theta}^T, \boldsymbol{\omega}^T)^T$, where $\boldsymbol{\omega} \equiv vech(\boldsymbol{\Omega})$. The pdf of the marginal distribution of $\mathbf{Y}_i$, denoted by $h(\mathbf{y}_i|\boldsymbol{\psi})$, is obtained by integrating out the unobserved random vector $\boldsymbol{\beta}_i$, i.e., $h(\mathbf{y}_i|\boldsymbol{\psi}) = \int g(\mathbf{y}_i|\boldsymbol{\beta}_i)f(\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Omega})\mathrm{d}\boldsymbol{\beta}_i$. Here and hereafter, integrations are taken over $\mathbb{R}^p$. Then, the log-likelihood function is defined as $l(\boldsymbol{\psi}) \equiv \sum \log h(\mathbf{y}_i|\boldsymbol{\psi})$. Hereafter, $h(\mathbf{y}_i|\boldsymbol{\psi})$, $g(\mathbf{y}_i|\boldsymbol{\beta}_i)$, and $f(\boldsymbol{\beta}_i|\boldsymbol{\theta}, \boldsymbol{\Omega})$ are abbreviated to h_i, g_i, and f_i, respectively, and the subscript i is suppressed where there is no ambiguity.

[2]Formally, the Dirac delta function is defined as the limit of a delta sequence and a normal pdf is its typical example.

[3]The proof of this result is available from the author upon request.

For a differentiable real-valued function $\phi(.)$ of an $m \times k$ matrix $\mathbf{A} = [a_{ij}]$, where m and k are positive integers, I denote the $m \times k$ matrix $[\partial\phi/\partial a_{ij}]$ by the symbol $\partial\phi/\partial\mathbf{A}$.

The first goal in this section is to derive a score test for the null hypothesis $\boldsymbol{\Omega} = \mathbf{0}$ (or $\boldsymbol{\omega} = \mathbf{0}$). The following Lemma plays an important role:

Lemma 1 *Let $f(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\Omega})$ be the pdf given in Assumption 1. Then*

$$\frac{\partial f}{\partial \boldsymbol{\Omega}} = \frac{1}{2}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\right] f(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\Omega}) = \frac{1}{2}\frac{\partial^2 f}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}. \tag{2}$$

From the above equation, we can see that $\frac{\partial f}{\partial \boldsymbol{\Omega}}$ is indeterminate under the null hypothesis $H_0 : \boldsymbol{\Omega} = \mathbf{0}$. To circumvent this difficulty, I define the score vector under the null hypothesis as the limit for which $\boldsymbol{\Omega} \to \mathbf{0}$.

When $\boldsymbol{\Omega} \to \mathbf{0}$, the limit of the pdf of a p-variate normal distribution is the p-dimensional *Dirac delta function*, which is denoted by $\delta_p(\boldsymbol{\beta}-\boldsymbol{\theta})$. The Dirac delta function can be viewed as a pdf with all the probability mass at a point. See, for example, Kanwal (1998) and Hoskins (1999) for details about the Dirac delta function, and see Peers (1971), Kobayashi (1991), Kobayashi and Shi (2005), and Kobayashi (2006) for a usage of Dirac delta function in statistics. The appendix of Kobayashi (2006) is useful for a quick review on properties of the Dirac delta function.

The p-dimensional Dirac delta function is known to satisfy

$$\text{(a)} \int g(\boldsymbol{\beta})\delta_p(\boldsymbol{\beta}-\boldsymbol{\theta})\mathrm{d}\boldsymbol{\beta} = g(\boldsymbol{\theta}), \qquad \text{(b)} \int g(\boldsymbol{\beta})\frac{\partial\delta_p}{\partial\boldsymbol{\beta}}\mathrm{d}\boldsymbol{\beta} = -\left.\frac{\partial g}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}}, \quad \text{and}$$
$$\text{(c)} \int g(\boldsymbol{\beta})\frac{\partial^2\delta_p}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\mathrm{d}\boldsymbol{\beta} = \left.\frac{\partial^2 g}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}},$$

for a sufficiently smooth function $g(\boldsymbol{\beta})$. Here $\frac{\partial\delta_p}{\partial\boldsymbol{\beta}}$ and $\frac{\partial^2\delta_p}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$ are obtained as the limits of $\frac{\partial f}{\partial\boldsymbol{\beta}}$ and $\frac{\partial^2 f}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$ for which $\boldsymbol{\Omega} \to \mathbf{0}$, respectively. Under the assumptions, we can differentiate under the integral sign, hence, the matrix of the first derivatives of $\log h = \log\int gf\mathrm{d}\boldsymbol{\beta}$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ are given, respectively, as $\frac{1}{h}\frac{\partial h}{\partial\boldsymbol{\theta}}$ or $\frac{1}{h}\int g\frac{\partial f}{\partial\boldsymbol{\theta}}\mathrm{d}\boldsymbol{\beta}$ and $\frac{1}{h}\frac{\partial h}{\partial\boldsymbol{\Omega}}$ or $\frac{1}{h}\int g\frac{\partial f}{\partial\boldsymbol{\Omega}}\mathrm{d}\boldsymbol{\beta}$. Since f is the p-dimensional Dirac delta function under the null hypothesis, and $\frac{\partial f}{\partial\boldsymbol{\theta}} = -\frac{\partial f}{\partial\boldsymbol{\beta}}$, we have, from (a), (b), and (c), that

$$\text{(a}'\text{)} \lim_{\boldsymbol{\Omega}\to\mathbf{0}} h = \lim_{\boldsymbol{\Omega}\to\mathbf{0}}\int gf\mathrm{d}\boldsymbol{\beta} = g|_{\boldsymbol{\beta}=\boldsymbol{\theta}}, \qquad \text{(b}'\text{)} \lim_{\boldsymbol{\Omega}\to\mathbf{0}}\int g\frac{\partial f}{\partial\boldsymbol{\theta}}\mathrm{d}\boldsymbol{\beta} = \left.\frac{\partial g}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}}, \quad \text{and}$$
$$\text{(c}'\text{)} \lim_{\boldsymbol{\Omega}\to\mathbf{0}}\int g\frac{\partial f}{\partial\boldsymbol{\Omega}}\mathrm{d}\boldsymbol{\beta} = \frac{1}{2}\lim_{\boldsymbol{\Omega}\to\mathbf{0}}\int g\frac{\partial^2 f}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\mathrm{d}\boldsymbol{\beta} = \frac{1}{2}\left.\frac{\partial^2 g}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}}.$$

Here, the first equality in (c$'$) comes from (2). Thus, the score vectors are given as $\mathbf{s}_{\boldsymbol{\theta}} \equiv \lim_{\boldsymbol{\Omega}\to\mathbf{0}}\frac{\partial l(\boldsymbol{\psi})}{\partial\boldsymbol{\theta}} = \sum\left[\frac{1}{g_i}\frac{\partial g_i}{\partial\boldsymbol{\beta}}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} = \sum\mathbf{d}_i$ and $\mathbf{s}_{\boldsymbol{\omega}} \equiv \lim_{\boldsymbol{\Omega}\to\mathbf{0}}\frac{\partial l(\boldsymbol{\psi})}{\partial\boldsymbol{\omega}} = \frac{1}{2}\sum vech\left[\frac{1}{g_i}\frac{\partial^2 g_i}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i^T}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}}$. Noting that $\frac{1}{g_i}\frac{\partial^2 g_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \frac{\partial\log g_i}{\partial\boldsymbol{\beta}}\frac{\partial\log g_i}{\partial\boldsymbol{\beta}^T} + \frac{\partial^2\log g_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$, we can write $\mathbf{s}_{\boldsymbol{\omega}} = \frac{1}{2}\sum vech\left[\frac{\partial\log g_i}{\partial\boldsymbol{\beta}_i}\frac{\partial\log g_i}{\partial\boldsymbol{\beta}_i^T} + \frac{\partial^2\log g_i}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i^T}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}}$ $= \frac{1}{2}\sum\mathbf{s}_i$. Chesher (1984) derived this equation with a different approach (under different assumptions). Finally, the score test statistic is defined as

$$\mathrm{ST} \equiv \mathbf{s}_{\boldsymbol{\omega}}^T(\widehat{\boldsymbol{\theta}})\;[\;\mathbf{R}\mathbf{Q}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{R}^T\;]\;\mathbf{s}_{\boldsymbol{\omega}}(\widehat{\boldsymbol{\theta}}), \tag{3}$$

where $\mathbf{R} \equiv [\;\mathbf{0}_{q\times p}\;\mathbf{I}_q]$, $\mathbf{Q} \equiv nE\begin{bmatrix}\mathbf{D}_i & \frac{1}{2}\mathbf{C}_i \\ \frac{1}{2}\mathbf{C}_i^T & \frac{1}{4}\mathbf{S}_i\end{bmatrix}$, $\mathbf{0}_{q\times p}$ is the $q \times p$ zero vector, and $\mathbf{I}_q$ is the $q \times q$ identity matrix. Here, $\widehat{\boldsymbol{\theta}}$ denotes the restricted MLE for the log-likelihood $l(\boldsymbol{\psi})$ under the restriction $\boldsymbol{\omega} = \mathbf{0}$, which is numerically the same as the MLE in (1), and so

I use the same notation. We can show that under the null hypothesis, the score test statistic follows asymptotically the chi-square distribution with $q \equiv p(p+1)/2$ degrees of freedom. See Davidson and MacKinnon (1993, p275) for a detailed description on the asymptotic properties of score tests.

In (3), $\mathbf{Q}(\widehat{\boldsymbol{\theta}})$ can be replaced by any consistent estimator. Two of the most commonly used estimators are the outer product of the gradient (OPG) estimator and the empirical Hessian (EH) estimator, which are defined, respectively, as

$$\text{(OPG estimator)} \quad \mathbf{Q}_{opg} \equiv \begin{bmatrix} \mathbf{D} & \frac{1}{2}\mathbf{C} \\ \frac{1}{2}\mathbf{C}^T & \frac{1}{4}\mathbf{S} \end{bmatrix}, \qquad \text{(EH estimator)} \quad \mathbf{Q}_{eh} \equiv -\mathbf{H},$$

where $\mathbf{H}$ is the Hessian matrix of the log likelihood with respect to the parameter vector $\boldsymbol{\psi}$ evaluated under the null hypothesis $\boldsymbol{\omega} = \mathbf{0}$. Let $\mathbf{H}$ be partitioned as $\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\boldsymbol{\theta\theta}} & \mathbf{H}_{\boldsymbol{\omega\theta}}^T \\ \mathbf{H}_{\boldsymbol{\omega\theta}} & \mathbf{H}_{\boldsymbol{\omega\omega}} \end{bmatrix}$, where, for example, $\mathbf{H}_{\boldsymbol{\theta\theta}}$ is the Hessian matrix with respect to $\boldsymbol{\theta}$, etc. Again by using the Dirac delta function approach and $\frac{\partial^2 f}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \frac{\partial^2 f}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$, we can easily show that $\mathbf{H}_{\boldsymbol{\theta\theta}} \equiv \lim_{\boldsymbol{\Omega}\to\mathbf{0}} \frac{\partial^2 l(\boldsymbol{\psi})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \sum \frac{\partial^2 \log g_i}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i^T}\Big|_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} = \mathbf{F}$. Elements of other submatrices, i,e, $\mathbf{H}_{\boldsymbol{\omega\theta}} \equiv \lim_{\boldsymbol{\Omega}\to\mathbf{0}} \frac{\partial^2 l(\boldsymbol{\psi})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\theta}^T}$ and $\mathbf{H}_{\boldsymbol{\omega\omega}} \equiv \lim_{\boldsymbol{\Omega}\to\mathbf{0}} \frac{\partial^2 l(\boldsymbol{\psi})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T}$ are given in the following proposition:

Proposition 1 *(Elements of the Hessian Matrix)*

$$\lim_{\boldsymbol{\Omega}\to\mathbf{0}} \frac{\partial^2 l(\boldsymbol{\psi})}{\partial\omega_{jk}\partial\theta_v} = \frac{1}{2}\sum \left[\frac{1}{g_i}\frac{\partial^3 g_i}{\partial\beta_j\partial\beta_k\partial\beta_v} - \frac{1}{g_i^2}\frac{\partial^2 g_i}{\partial\beta_j\partial\beta_k}\frac{\partial g_i}{\partial\beta_v}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}}, \tag{4}$$

$$\lim_{\boldsymbol{\Omega}\to\mathbf{0}} \frac{\partial^2 l(\boldsymbol{\psi})}{\partial\omega_{jk}\partial\omega_{vw}} = \frac{1}{4}\sum \left[\frac{1}{g_i}\frac{\partial^4 g_i}{\partial\beta_j\partial\beta_k\partial\beta_v\partial\beta_w}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} - \frac{1}{4}\sum \left[\frac{1}{g_i^2}\frac{\partial^2 g_i}{\partial\beta_j\partial\beta_k}\frac{\partial^2 g_i}{\partial\beta_v\partial\beta_w}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}}, \tag{5}$$

for j, k, v, w =1,...,p. Here, for example, I abbreviate β_{ik} to β_k etc.

From the identity $\frac{1}{g_i}\frac{\partial^3 g_i}{\partial\beta_j\partial\beta_k\partial\beta_v} - \frac{1}{g_i^2}\frac{\partial^2 g_i}{\partial\beta_j\partial\beta_k}\frac{\partial g_i}{\partial\beta_v} = \frac{\partial^3 \log g_i}{\partial\beta_j\partial\beta_k\partial\beta_v} + \frac{\partial^2 \log g_i}{\partial\beta_j\partial\beta_v}\frac{\partial \log g_i}{\partial\beta_k} + \frac{\partial^2 \log g_i}{\partial\beta_k\partial\beta_v}\frac{\partial \log g_i}{\partial\beta_j}$, we can see that the right-hand side in (4) is an element of $\frac{1}{2}\mathbf{G}$, and thus we have $\mathbf{H}_{\boldsymbol{\omega\theta}} = \frac{1}{2}\mathbf{G}$. Noting that the second term on the right-hand side in (5) is an element of $\frac{1}{4}\mathbf{S}$, we can see that $\mathbf{H}_{\boldsymbol{\omega\omega}} = \frac{1}{4}\left(\mathbf{A} - \mathbf{S}\right)$, where

$$\underset{q\times q}{\mathbf{A}} \equiv \begin{bmatrix} \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_1)^4}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} & \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_1)^3\partial\beta_2}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} & \cdots & \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_1)^2\partial(\beta_p)^2}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} \\ \vdots & \vdots & & \vdots \\ \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_p)^2(\partial\beta_1)^2}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} & \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_p)^2\partial\beta_1\partial\beta_2}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} & \cdots & \sum\left[\frac{1}{g_i}\frac{\partial^4 g_i}{(\partial\beta_p)^4}\right]_{\boldsymbol{\beta}_i=\boldsymbol{\theta}} \end{bmatrix}.$$

From these arguments, we have

$$\mathbf{Q}_{eh} = \begin{bmatrix} -\mathbf{F} & -\frac{1}{2}\mathbf{G}^T \\ -\frac{1}{2}\mathbf{G} & \frac{1}{4}(\mathbf{S} - \mathbf{A}) \end{bmatrix}.$$

This EH estimator for the score test has not been derived in the literature and is used to show the equivalence between a specification robust form of the score test and the IM test with White's ACME in the next subsection.

2.3 Specification Robust Form of the Score Test

Let $\boldsymbol{\Psi}$ be the parameter space of $\boldsymbol{\psi}$. In the previous subsection, I supposed that the true model belongs to the known model structure $h(\mathbf{y}_i|\boldsymbol{\psi})$, i.e., the true pdf is given by $h(\mathbf{y}_i|\boldsymbol{\psi}_0)$ for some $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$; the proposed score test examines whether $\boldsymbol{\psi}_0 = (\boldsymbol{\theta}_0^T, \boldsymbol{\omega}_0^T)^T = (\boldsymbol{\theta}_0^T, \mathbf{0}^T)^T$.

In practice, this assumption is rarely true. If the true pdf is not given by $h(\mathbf{y}_i|\boldsymbol{\psi})$ for any $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, that is, the model is misspecified, then the MLE $\widehat{\boldsymbol{\psi}}$, or a solution that maximizes the misspecified log likelihood, is called a "quasi-maximum likelihood estimator" (QMLE). A QMLE may be consistent for a parameter of interest. For example, for a classical linear regression model, the QMLE for the coefficient vector under Gaussian error assumption is consistent even when the true distribution of the errors is not Gaussian. White (1982, p.4) showed that the QMLE is strongly consistent for the parameter vector which minimizes the Kullback Leibler Information Criterion (KLIC) (see White, 1982, p.3 for the definition). Intuitively the KLIC measures our ignorance about the true model structure, and hence White (1982) called the QMLE "minimum ignorance estimator."

Let $\boldsymbol{\psi}_* \equiv (\boldsymbol{\theta}_*^T, \boldsymbol{\omega}_*^T)^T$ denote the parameter vector which minimizes the KLIC for the misspecified log-likelihood $l(\boldsymbol{\psi}) = \sum \log h(\mathbf{y}_i|\boldsymbol{\psi})$ (if the model is correctly specified, $\boldsymbol{\psi}_* = \boldsymbol{\psi}_0$). We may wish to test whether $\boldsymbol{\psi}_*$ satisfies a certain constrain; however, the asymptotic distribution of the usual form of score tests under model misspecification is different from that under correct model specification. White (1982) proposed a robust form of score tests applicable to QMLEs for testing the null hypothesis $H_0 : w(\boldsymbol{\psi}_*) = \mathbf{0}$, where $w : \mathbb{R}^m \to \mathbb{R}^r$ is a continuous vector function of $\boldsymbol{\psi}_*$ such that its Jacobian at $\boldsymbol{\psi}_*$ is finite with full row rank r, against the alternative $H_1 : w(\boldsymbol{\psi}_*) \neq \mathbf{0}$. In the present case, $H_0 : \boldsymbol{\omega}_* = \mathbf{0}$ and the robust form of the score test is given as

$$\mathrm{ST}^* \equiv \mathbf{s}_{\boldsymbol{\omega}}^T(\widehat{\boldsymbol{\theta}}) \left[\, \mathbf{R}\mathbf{Q}_{eh}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{R}^T \,\right] \left[\, \mathbf{R}\mathbf{Q}_{sw}(\widehat{\boldsymbol{\theta}})\mathbf{R}^T \,\right]^{-1} \left[\, \mathbf{R}\mathbf{Q}_{eh}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{R}^T \,\right] \mathbf{s}_{\boldsymbol{\omega}}(\widehat{\boldsymbol{\theta}}), \tag{6}$$

where $\widehat{\boldsymbol{\theta}}$ is the restricted QMLE, i.e., a solution that maximizes the misspecified restricted log-likelihood $l(\boldsymbol{\theta}) = \sum \log g(\mathbf{y}_i|\boldsymbol{\theta})$, and $\mathbf{Q}_{sw}$ is a so-called "sandwich estimator", which is defined as

$$\text{(Sandwich estimator)} \quad \mathbf{Q}_{sw} \equiv \mathbf{Q}_{eh}^{-1}\mathbf{Q}_{opg}\mathbf{Q}_{eh}^{-1}.$$

It can be shown that ST* asymptotically follows $\chi^2(q)$ under the null hypothesis $\boldsymbol{\omega}_* = \mathbf{0}$ (White, 1982, Theorem 3.5). The following Proposition is the main result in this note:

Proposition 2 *The two test statistics, IM and ST*, defined in (1) and (6), respectively, are numerically the same if we use the same estimator for $\boldsymbol{\theta}$.*

Proposition 2 shows that the IM test with White's ACME is equivalent to the specification robust form of the score test. Recall that the information matrix equality is merely a necessary condition for a model to be correctly specified. If $E[\mathbf{D}_i(\boldsymbol{\theta}_*) + \mathbf{F}_i(\boldsymbol{\theta}_*)] = \mathbf{0}$ holds, then the IM test with White's ACME is totally powerless against this situation even though the model is misspecified. On the other hand, if our objective is to compare two models, constant parameter and random parameter models, and both models possibly be misspecified models, we should use the specification robust form of the score test.

3. Concluding Remarks

Recently Gan and Jiang (1999) have given an interesting interpretation to the IM test; the IM test examines whether a root of the likelihood equation corresponds with a global maximum. The assumption that the random parameter vectors are independent can be relaxed so that the random parameter vector follows a stochastic process, and a score test for this case would be derived by a similar approach. See Kobayashi and Shi (2005) for a usage of Dirac delta function in this context for specific cases.

Appendix

Proof of Lemma 1: The logarithm of f is given by $l_f(\boldsymbol{\beta}|\boldsymbol{\theta},\boldsymbol{\Omega}) = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega})$,where $q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega}) \equiv (\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})$. Noting that trace$[q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega})] = q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega})$ (since $q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega})$ is a scalar), the derivative of $q(\boldsymbol{\beta};\boldsymbol{\theta},\boldsymbol{\Omega})$ with respect to (i,j) element of $\boldsymbol{\Omega}$, i.e., ω_{ij}, is

$$\begin{aligned}\frac{\partial q}{\partial \omega_{ij}} &= \text{trace}\left[(\boldsymbol{\beta}-\boldsymbol{\theta})^T\tfrac{\partial \boldsymbol{\Omega}^{-1}}{\partial \omega_{ij}}(\boldsymbol{\beta}-\boldsymbol{\theta})\right] \\ &= -\text{trace}\left[(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}\tfrac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})\right] \\ &= -\text{trace}\left[\tfrac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}\right] \\ &= -m_{ij},\end{aligned} \tag{7}$$

where m_{ij} is (i,j) element of the matrix $\mathbf{M} \equiv \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}$. The second and third equalities are obtained by using that $\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \omega_{ij}} = -\boldsymbol{\Omega}^{-1}(\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}})\boldsymbol{\Omega}^{-1}$ and trace$(\mathbf{ABC}) =$ trace$(\mathbf{BCA})$ for conformable matrices, $\mathbf{A}$,$\mathbf{B}$, and $\mathbf{C}$. Note that $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}$ is the $p \times p$ matrix whose (i,j) element is one and all other elements are zero. Pre-multiplying $\mathbf{M}$ by $\frac{\partial \boldsymbol{\Omega}}{\partial \omega_{ij}}$ and taking its trace leads to (i,j) element of $\mathbf{M}$. From (7), we have

$$\frac{\partial q}{\partial \boldsymbol{\Omega}} = \begin{bmatrix} -m_{11} & \cdots & -m_{1p} \\ \vdots & \ddots & \vdots \\ -m_{p1} & \cdots & -m_{pp} \end{bmatrix} = -\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1}. \tag{8}$$

From (8) and $\frac{\partial \log|\boldsymbol{\Omega}|}{\partial \boldsymbol{\Omega}} = \boldsymbol{\Omega}^{-1}$, we have $\frac{\partial l_f}{\partial \boldsymbol{\Omega}} = \frac{1}{2}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\right]$. From this equation and the identity, $\frac{\partial f}{\partial \boldsymbol{\Omega}} = f(\frac{\partial l_f}{\partial \boldsymbol{\Omega}})$, the first equality in (2) follows. By a standard matrix calculus (see, e.g., Magnus and Neudecker, 1999) we can show that $\frac{\partial l_f}{\partial \boldsymbol{\beta}} = -\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}-\boldsymbol{\theta})$, and $\frac{\partial^2 l_f}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = -\boldsymbol{\Omega}^{-1}$, from which and the identity, $\frac{\partial^2 f}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = f\left[\frac{\partial l_f}{\partial \boldsymbol{\beta}}\frac{\partial l_f}{\partial \boldsymbol{\beta}^T} + \frac{\partial^2 l_f}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\right]$, we have the second equality in (2), which completes the proof of Lemma 1. □

Proof of Proposition 1: Since $h = \int gf \mathrm{d}\boldsymbol{\beta}$, we have (A) $\frac{\partial \log h}{\partial \omega_{jk}} = \frac{1}{h}\frac{\partial h}{\partial \omega_{jk}} = \frac{1}{h}\int g\frac{\partial f}{\partial \omega_{jk}}\mathrm{d}\boldsymbol{\beta}$. Differentiating the most left and the most right hand sides of this equation with respect to ω_{vw}, we have (B) $\frac{\partial^2 \log h}{\partial \omega_{jk}\partial \omega_{vw}} = -\frac{1}{h^2}\frac{\partial h}{\partial \omega_{vw}}\int g\frac{\partial f}{\partial \omega_{jk}}\mathrm{d}\boldsymbol{\beta} + \frac{1}{h}\int g\frac{\partial^2 f}{\partial \omega_{jk}\partial \omega_{vw}}\mathrm{d}\boldsymbol{\beta}$. From Lemma 1, we have $\frac{\partial^2 f}{\partial \beta_j \partial \beta_k} = 2\frac{\partial f}{\partial \omega_{jk}}$, and thus (C) $\frac{\partial^4 f}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w} = \frac{\partial^2}{\partial \beta_v\partial \beta_w}\left[2\frac{\partial f}{\partial \omega_{jk}}\right] = 2\frac{\partial}{\partial \omega_{jk}}\left[\frac{\partial^2 f}{\partial \beta_v\partial \beta_w}\right] = 4\frac{\partial^2 f}{\partial \omega_{jk}\partial \omega_{vw}}$. From (A), (B), and (C), we have

$$\frac{\partial^2 \log h}{\partial \omega_{jk}\partial \omega_{vw}} = -\left(\frac{1}{h}\frac{\partial h}{\partial \omega_{jk}}\right)\left(\frac{1}{h}\frac{\partial h}{\partial \omega_{vw}}\right) + \frac{1}{4h}\int_{\mathbb{R}^p} g\frac{\partial^4 f}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w}\mathrm{d}\boldsymbol{\beta}. \tag{9}$$

Under the null hypothesis $\boldsymbol{\Omega} \to \mathbf{0}$, f is the p dimensional Dirac delta function, which satisfies $\int g(\boldsymbol{\beta})\frac{\partial^4 \delta^p(\boldsymbol{\beta}-\boldsymbol{\theta})}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w}\mathrm{d}\boldsymbol{\beta} = \left.\frac{\partial^4 g(\boldsymbol{\beta})}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}}$ for a sufficiently smooth function $g(\boldsymbol{\beta})$ and for j, k, v, w =1, ...p. Thus, under the null hypothesis and Assumption 2, the integral in (9) reduces to $\left.\frac{\partial^4 g}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w}\right|_{\boldsymbol{\beta}=\boldsymbol{\theta}}$. Hence, we have

$$\lim_{\boldsymbol{\Omega}\to\mathbf{0}}\frac{\partial \log h}{\partial \omega_{jk}\partial \omega_{vw}} = \left[-\frac{1}{4g^2}\frac{\partial^2 g}{\partial \beta_j\partial \beta_k}\frac{\partial^2 g}{\partial \beta_v\partial \beta_w} + \frac{1}{4g}\frac{\partial^4 g}{\partial \beta_j\partial \beta_k\partial \beta_v\partial \beta_w}\right]_{\boldsymbol{\beta}=\boldsymbol{\theta}}.$$

Equation (5) in Proposition 1 follows from this equation. The first equation in Proposition 1 can be obtained in the same fashion, thus the proof is omitted. □

Proof of Proposition 2: I suppose that the same estimator for $\boldsymbol{\theta}$ is used in calculating both statistics, and hence the argument $\boldsymbol{\theta}$ is suppressed throughout the proof. Let $\mathbf{Q}_{eh}^{-1}$ be partitioned as $\mathbf{Q}_{eh}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix}$. Note that $\mathbf{B}_{22} = \mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T$. After a simple calculation, we can show that

$$\begin{aligned}\mathbf{R}\mathbf{Q}_{sw}\mathbf{R}^T &= \mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{Q}_{opg}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T \\ &= \mathbf{B}_{12}^T\mathbf{D}\mathbf{B}_{12} + \tfrac{1}{2}\mathbf{B}_{12}^T\mathbf{C}\mathbf{B}_{22} + \tfrac{1}{2}\mathbf{B}_{22}\mathbf{C}\mathbf{B}_{12} + \tfrac{1}{4}\mathbf{B}_{22}\mathbf{S}\mathbf{B}_{22}.\end{aligned}$$

Applying the usual matrix inversion formula for partitioned matrices, it is easy to show $\mathbf{B}_{12}^T = -\frac{1}{2}\mathbf{B}_{22}\mathbf{G}\mathbf{F}^{-1}$, and so $\mathbf{B}_{22}^{-1}\mathbf{B}_{12}^T = -\frac{1}{2}\mathbf{G}\mathbf{F}^{-1}$. Hence, we have

$$\begin{aligned}&\left(\mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T\right)^{-1}\left(\mathbf{R}\mathbf{Q}_{sw}\mathbf{R}^T\right)\left(\mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T\right)^{-1} \\ &= \mathbf{B}_{22}^{-1}\left[\mathbf{B}_{12}^T\mathbf{D}\mathbf{B}_{12} + \tfrac{1}{2}\mathbf{B}_{12}^T\mathbf{C}\mathbf{B}_{22} + \tfrac{1}{2}\mathbf{B}_{22}\mathbf{C}\mathbf{B}_{12} + \tfrac{1}{4}\mathbf{B}_{22}\mathbf{S}\mathbf{B}_{22}\right]\mathbf{B}_{22}^{-1} \\ &= \mathbf{B}_{22}^{-1}\mathbf{B}_{12}^T\mathbf{D}\mathbf{B}_{12}\mathbf{B}_{22}^{-1} + \tfrac{1}{2}\mathbf{B}_{22}^{-1}\mathbf{B}_{12}^T\mathbf{C} + \tfrac{1}{2}\mathbf{C}\mathbf{B}_{12}\mathbf{B}_{22}^{-1} + \tfrac{1}{4}\mathbf{S} \\ &= \tfrac{1}{4}\left[\mathbf{G}\mathbf{F}^{-1}\mathbf{D}\mathbf{F}^{-1}\mathbf{G}^T - \mathbf{G}\mathbf{F}^{-1}\mathbf{C} - \mathbf{C}^T\mathbf{F}^{-1}\mathbf{G}^T + \mathbf{S}\right] \\ &= \tfrac{1}{4}\mathbf{Q}_W,\end{aligned}$$

from which and $\mathbf{s}_{\boldsymbol{\omega}} = \mathbf{s}/2$, we have

$$\begin{aligned}\mathrm{ST}^* &= \mathbf{s}_{\boldsymbol{\omega}}^T\left[\left(\mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T\right)^{-1}\left(\mathbf{R}\mathbf{Q}_{sw}\mathbf{R}^T\right)\left(\mathbf{R}\mathbf{Q}_{eh}^{-1}\mathbf{R}^T\right)^{-1}\right]^{-1}\mathbf{s}_{\boldsymbol{\omega}} \\ &= \mathbf{s}^T\mathbf{Q}_W^{-1}\mathbf{s}. \qquad \square\end{aligned}$$

References

Chesher, A. D. (1983) "The information matrix test: simplified calculation via a score test interpretation" *Economics Letters* **13**, 45-48.

Chesher, A. D. (1984) "Testing for neglected heterogeneity" *Econometrica* **52(4)**, 865-872.

Davidson, R., and MacKinnon, J. G. (1993) *Estimation and Inference in Econometrics*, Oxford University Press: New York.

Dhaene, G., and Hoorelbeke, D. (2004) "The information matrix test with bootstrap-based covariance matrix estimation" *Economics Letters* **82**, 341-347.

Gan, L., and Jiang, J. (1999) "A test for global maximum" *Journal of the American Statistical Association* **94**, 847-854.

Kanwal, R. P. (1998) *Generalized Functions: Theory and Technique* 2nd ed., Birkhauser: Boston.

Kobayashi, M. (1991) "Testing for autocorrelated disturbances in nonlinear regression analysis" *Econometrica* **59(4)**, 1153-1159.

Kobayashi, M. (2006) "Testing for volatility jumps in the stochastic volatility process" *Asia-Pacific Financial Markets* **12**, 143-157.

Kobayashi, M., and Shi X.-H. (2005) "Testing for EGARCH against stochastic volatility models" *Journal of Time Series Analysis* **26(1)**, 135-150.

Orme, C. (1990) "The small sample performance of the information matrix test" *Journal of Econometrics* **46**, 309-331.

Horowitz, J. L. (1994) "Bootstrap-based critical values for the information matrix test" *Journal of Econometrics* **61**, 395-411.

Hoskins, R. F. (1999) *Delta Functions: An Introduction to Generalized Functions*, Horwood Publishing Limited: Sussex, England.

Lancaster, T. (1984) "The covariance matrix of the information matrix test" *Econometrica* **52(4)**, 1051-1053.

Magnus, J. R., and Neudecker, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics* revised ed., John Wiley and Sons: New York.

Peers, W. H. (1971) "Likelihood ratio and associated test criteria" *Biometrika* **58(3)**, 577-587.

White, H. (1982) "Maximum likelihood estimation of misspecified models" *Econometrica* **50(1)**, 1-25.

White, H. (1983) Corrigendum, *Econometrica* **51(2)**, 513.